SYNOPSIS

DNA binding studies with the transcriptional activator protein C of bacteriophage *Mu*

The family of DNA binding proteins from prokaryotes were amongst the earliest to be studied and characterized at the molecular level They serve today as paradigms for understanding the principles involved in DNA protein recognition The interaction between DNA and the regulatory proteins is fundamental since it influences a crucial step in gene expression i e transcription initiation Regulation of gene expression in prokaryotes is subjected to both positive and negative control The genes which are subjected to positive control are expressed only when they are activated' by the regulatory protein Activation of transcription by these proteins is known to occur either by direct contact with RNA polymerase and/or by the activator mediated DNA distortion A critical step in both the mechanisms is the binding and recognition of DNA by the regulatory protein at its cognate site The structural and chemical complementarity between the interacting surfaces on the macromolecules is an important determinant in this recognition process which is characterized by a high binding affinity and sequence specificity

The positive and negative control circuits which operate to modulate the expression of a gene are not mutually exclusive, a classical example being the *lac* operon, which is subjected to both forms of regulation Each system has evolved its own mode of control and as more and more systems are studied, the intricacies involved in their regulatory mechanisms become evident The *mom* gene of bacteriophage *Mu* is a case which highlights this point very well This gene, which encodes a unique DNA modification function, is the last gene to be expressed in the lytic cycle of the phage, and is subjected to a complex mode of regulation The expression of *mom* gene is dependent on the action of two host proteins, the *E coli* Dam methylase and the repressor OxyR The methylation dependent expression of this gene was the first example demonstrating the involvement of methylation in positive control The *mom* gene expression is further modulated by two phage encoded proteins C and Com The C protein, a middle gene

product, positively regulates transcription, while the late gene product Com stimulates the translation of *mom* mRNA The balance and interplay amongst these factors acting at various levels, decide the amounts and temporal fate of *mom* gene product **This thesis work has been undertaken to investigate the mode of interaction of the transcriptional activator protein C with DNA, so as to understand the mechanism of C-mediated transcription activation** The following features of C protein relevant to its DNA binding function have been characterized

1 the nature of DNA binding with respect to specificity and affinity

2 oligomeric status of the protein in its native state and when bound to DNA

3 the region in the protein involved in DNA binding

4 nucleotide contact points located within the DNA binding site

Chapter 1 is a general introduction to the control circuits in gene regulation with emphasis on positive control mechanisms in transcription activation The mode of interaction and activation of transcription by the *E coli* cyclic AMP receptor protein and the bacteriophage λcI have been reviewed as these two proteins are probably the best understood gene regulatory proteins involved in positive control In addition two other regulatory proteins, MerR and FadR have been discussed, which present certain unique features in their mechanism of transcription activation A brief account on the well characterized DNA binding motifs and on principles of sequence specific DNA recognition is given The chapter ends with an introduction to bacteriophage *Mu* and on the regulation of *mom* gene expression of the phage with emphasis on the role of C protein which is the topic of this investigation

To achieve the objectives of this study it was desirable to construct a clone which overexpresses the C protein so as to aid in its purification The work presented in **chapter 2** demonstrates the involvement of a secondary structure at the translation initiation region of the *C* mRNA which caused low levels of protein production Removal of the sequence contributing to this structure resulted in hyperexpression of the gene The presence of a structure in RNA was also demonstrated by electrophoretic analysis of short *in vitro* transcripts comprising the

sequences which contribute to its formation

The experimental approach involved in the construction of a clone overproducing the C protein yielded a recombinant protein which does not have second and third amino acids, Q and H respectively To assay its activity the site-specific DNA binding property of C protein was used Crude extracts from expressing clones showed a site-specific DNA binding activity when a labelled DNA fragment containing the C binding site was used in the assays To further characterize this activity the recombinant overexpressed protein was purified by specific affinity chromatographic techniques The results on these lines form the basis of **chapter 3** In addition, an attempt has been made to probe the molecular and thermodynamic origins of stability and specificity by studying the equilibrium binding properties of the purified protein as a function of solution conditions

Site-specific DNA binding proteins predominantly recognize dyad symmetric sequences and bind to these sites with two or four identical subunits In the crystal structure of many DNA protein complexes, the axis of two-fold symmetry of the protein is coincident with that of the two-fold symmetric recognition site Dimerisation can occur prior to or during the process of DNA binding The binding site for C protein located in the promoter region of *mom* includes an interrupted palindromic sequence The experiments presented in **chapter 4** have addressed the question of oligomeric status of C protein in its native state and when bound to DNA The results indicate that the protein exists in solution predominantly as a dimer and also binds DNA as a dimer Further, the primary sequence of C protein shows the presence of certain putative DNA binding motifs The results on deletion analysis in the *C* gene implicate a role for the C-terminal region of the protein, which has a putative helix-turn-helix motif, in DNA binding

In **Chapter 5**, the results of footprinting experiments performed to map the probable contact points within the C binding site are presented In DNaseI protection experiments a 28 bp region was protected by the bound C protein compared to the 22-24 bp normally seen for proteins whose sizes are comparable to the dimeric C Dimethyl sulfate mapped the purine contacts and located the interaction of the protein at two adjacent major grooves Further, the

guanines positioned at periodic intervals, in the bottom strand, showed hyper-reactivity upon protein binding Interference footprinting with ethyl nitrosourea located the phosphate contacts and hydroxyl radical [Fe(II)-EDTA] protection revealed a tripartite footprint on the same face of the helix Taken together, these results indicate that a dimer of C protein, interacting across the minor groove, binds at the center and makes contact beyond the palindromic sequence with two adjacent major grooves Finally, the chemical nuclease 1,10-phenanthroline-copper, a minor groove specific ligand, showed hyper-reactivity upon protein binding, in the top strand A summary of all the footprinting data is shown in Figure 1A

The *mom* promoter has two important features to qualify as an activator dependent weak promoter a poor -35 element (ACCACA) compared to the canonical sequence (TTGACA) and a sub-optimal 19 bp spacing between the -10 and -35 elements The C binding site in this promoter is located upstream and overlapping the -35 region (Figure 1B) Two features of C binding are striking the large span of the protected region and the periodic location of hyper-reactive guanines One of the possible interpretations of this data could be a change in DNA conformation resulting in its bending towards the protein Further, the hyper-reactivity seen in the nucleotide triplet CAC overlapping the -35 element includes the TGTG motif which is prone to helical deformations The enhanced reactivity upon protein binding detected by a minor groove specific ligand could be a consequence of widened minor groove Widening of minor groove alters the torsional flexibility of DNA and leads to localized unwinding The 19 bp spacing between the -10 and -35 elements of P_{mom}, compared to the 17±1 bp seen in most promoters, means an additional twist angle of atleast 34^{o} The C protein induced localized unwinding of DNA can compensate for this amount of twist angle difference and could be a crucial factor in the mechanism of transcription activation In addition to this phenomenon, the C protein might also have contacts with RNAP and influence activation of transcription

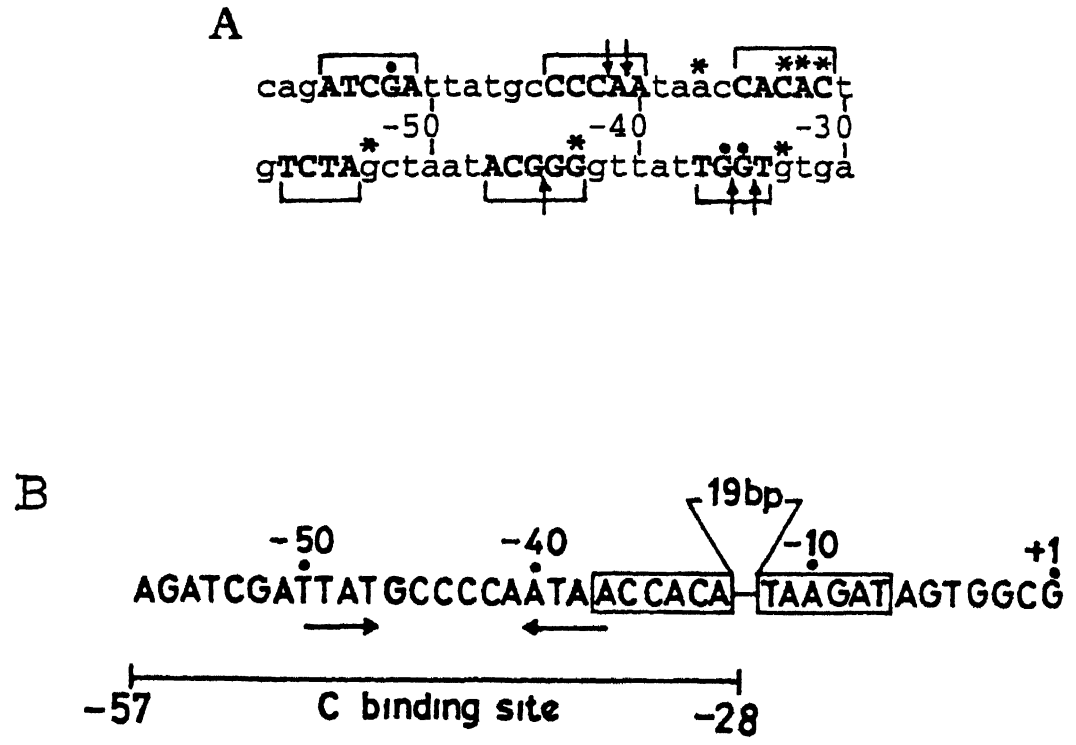

Figure 1 **A** Summary of footprinting data The hydroxyl radical protections are in uppercase and bracketed Hyper-reactive residues are marked with asterisks and the protected guanines are indicated by filled circles The vertical arrows indicate the position of phosphate interference

B The *mom* regulatory region showing the location of C binding site The -10 and -35 promoter elements are boxed The dyad symmetric sequence within the C binding site are indicated by arrows